Revisiting the Neyman-Scott model: an Inconsistent MLE or an Ill-defined Model?

Aris Spanos
Department of Economics,
Virginia Tech, USA

January 2013

Abstract

The Neyman and Scott (1948) model is widely used to demonstrate a serious weakness of the Maximum Likelihood (ML) method: it can give rise to inconsistent estimators. The primary objective of this paper is to revisit this example with a view to demonstrate that the culprit for the inconsistent estimation is not the ML method but an ill-defined statistical model. It is also shown that a simple recasting of this model renders it well-defined and the ML method gives rise to consistent and asymptotically efficient estimators.

1 Introduction

Despite claims for priority by a number of different authors, Maximum Likelihood (ML) was first articulated as a general method of estimation in the context of a parametric statistical model by Fisher (1922). It took several decades to establish the regularity conditions needed to ensure the key asymptotic properties of ML Estimators (MLE), such as efficiency and consistency (Cramer, 1946, Wald, 1949), but since then ML has dominated estimation in frequentist statistics; see Stigler (2007), Hald (2007) for this history.

Several counterexamples were proposed in the 1940s and 1950s raising doubts about the generality of the ML method. These counterexamples include Hodges's local superefficient estimator (Le Cam, 1953), the mixture of two Normal distributions (Cox, 2006)) and the inconsistent MLE example proposed by the Neyman and Scott (1948) model. Commenting on these examples Stigler (2007), p. 613, argued that none are considered serious enough to undermine the credibility of the ML method, and singled out the last example:

"The Wald-Neyman-Scott example was of more practical import, and still serves as a warning of what might occur in modern highly parameterized problems, where the information in the data may be spread too thinly to achieve asymptotic consistency."

The primary objective of this note is to revisit the Neyman-Scott example with a view to unpack Stigler's assessment by demonstrating that the real culprit for the inconsistent estimator is not the ML method, as such, but an ill-defined statistical model. It is also shown that a simple recasting of this model renders it well-defined and the ML method gives rise to consistent and asymptotically efficient estimators.

2 The Neyman-Scott model

The quintessential example used to demonstrate that ML might give rise to inconsistent estimators is *the Neyman-Scott model*:

$$\boxed{\left.\begin{array}{c} X_{it}=\mu_t+\varepsilon_{it}, \\ \varepsilon_{it}\backsim\mathsf{NIID}(0,\sigma^2)=0, \end{array}\right\}\ i{=}1,2,\ t{=}1,2,...,n,...}$$

which can be viewed as a simple time effects panel data model.

The underlying distribution of the observable random variables (X_{1t}, X_{2t}) is bivariate Normal Independent, but *not* Identically Distributed:

$$\mathbf{X}_t{:=}\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix}\backsim\mathsf{NI}\left(\begin{pmatrix} \mu_t \\ \mu_t \end{pmatrix},\begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}\right),\ t{=}1,2,...,n,...$$

In light of the fact that the non-ID assumption implies that this model suffers from the *incidental parameter problem*, in the sense that the unknown parameters $(\mu_1, \mu_2, ..., \mu_n)$ increase with the sample size, the latter are viewed as *nuisance* parameters, and σ^2 as the only parameter of interest.

3 Maximum Likelihood Estimation?

Despite the incidental parameter problem the Neyman-Scott model continues to be used as a counterexample to the ML method. That is, the literature ignores the incidental parameter problem, and defines the distribution of the sample to be:

$$f(\mathbf{x};\boldsymbol{\theta}) = \prod_{t=1}^{n}\prod_{i=1}^{2}\frac{1}{\sigma\sqrt{2\pi}}e^{\left\{-\frac{1}{2\sigma^2}(x_{it}-\mu_t)^2\right\}} = \prod_{t=1}^{n}\frac{1}{2\pi\sigma^2}e^{\left\{-\frac{1}{2\sigma^2}[(x_{1t}-\mu_t)^2+(x_{2t}-\mu_t)^2]\right\}}$$

This gives rise to the 'log-likelihood function':

$$\ln L(\boldsymbol{\theta};\mathbf{x}) = -\,n\ln\sigma^2 - \tfrac{1}{2\sigma^2}\textstyle\sum_{t=1}^{n}[(x_{1t}-\mu_t)^2+(x_{2t}-\mu_t)^2].$$

The 'Maximum Likelihood Estimators' (MLE) are supposed to be derived by solving the first-order conditions:

$$\frac{\partial\ln L(\boldsymbol{\theta};\mathbf{x})}{\partial\mu_t} = \tfrac{1}{\sigma^2}[(x_{1t}-\mu_t)+(x_{2t}-\mu_t)]=0,$$

$$\frac{\partial\ln L(\boldsymbol{\theta};\mathbf{x})}{\partial\sigma^2} = -\tfrac{n}{\sigma^2}+\tfrac{1}{2\sigma^4}\textstyle\sum_{t=1}^{n}[(x_{1t}-\mu_t)^2+(x_{2t}-\mu_t)^2]=0,$$

giving rise to:

$$\widehat{\mu}_t=\tfrac{1}{2}(X_{1t}+X_{2t}),\ \ t=1,2,...,n,$$

$$\widehat{\sigma}^2=\frac{1}{2n}\sum_{t=1}^{n}[(X_{1t}-\widehat{\mu}_t)^2+(X_{2t}-\widehat{\mu}_t)^2]=\frac{1}{n}\sum_{t=1}^{n}s_t^2,$$

where $s_t^2=\frac{1}{2}[(X_{1t}-\widehat{\mu}_t)^2+(X_{2t}-\widehat{\mu}_t)^2]$.

Notice that for $\ln L(\boldsymbol{\theta};\mathbf{x})$, $\widehat{\boldsymbol{\theta}}_{MLE}:=(\widehat{\mu}_t,\hat{\sigma}^2,t=1,2,...,n)$ is a maximum since the second derivatives at $\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}$ are:

$$\left.\frac{\partial^2\ln L(\boldsymbol{\theta};\mathbf{x})}{\partial\mu_t^2}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_{MLE}} = -\left(\tfrac{2}{\sigma^2}\right)\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_{MLE}} = -\tfrac{2}{\hat{\sigma}^2}<0,$$

$$\left.\frac{\partial^2\ln L(\boldsymbol{\theta};\mathbf{x})}{\partial\sigma^2\partial\mu_t}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_{MLE}} = -\tfrac{1}{\sigma^4}\textstyle\sum_{t=1}^{n}[(x_{1t}-\mu_t)+(x_{2t}-\mu_t)]\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_{MLE}}=0$$

$$\left.\frac{\partial^2\ln L(\boldsymbol{\theta};\mathbf{x})}{\partial\sigma^4}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_{MLE}} = \tfrac{n}{\sigma^4}-\tfrac{1}{\sigma^6}\textstyle\sum_{t=1}^{n}[(x_{1t}-\mu_t)^2+(x_{2t}-\mu_t)^2]\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_{MLE}} = -\tfrac{n}{\hat{\sigma}^4}<0,$$

and thus:

$$\left.\left(\frac{\partial^2\ln L(\boldsymbol{\theta};\mathbf{x})}{\partial\mu_t^2}\right)\left(\frac{\partial^2\ln L(\boldsymbol{\theta};\mathbf{x})}{\partial\sigma^4}\right)-\left(\frac{\partial^2\ln L(\boldsymbol{\theta};\mathbf{x})}{\partial\sigma^2\partial\mu_t}\right)\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_{MLE}}>0.$$

The commonly used argument against the ML method is that since:

$$E(\widehat{\mu}_t)=\mu_t \text{ and } E(s_t^2)=\tfrac{1}{2}\sigma^2,$$

it follows that the 'MLE' $\widehat{\sigma}^2$ is both *biased* and *inconsistent* because:

$$E(\widehat{\sigma}^2)=\tfrac{1}{n}\textstyle\sum_{t=1}^{n}E(s_t^2)=\tfrac{1}{2}\sigma^2, \quad \text{and} \quad \widehat{\sigma}^2 \overset{a.s.}{\to} \tfrac{1}{2}\sigma^2,$$

since the bias $E(\widehat{\sigma}^2)-\sigma^2=-\frac{1}{2}\sigma^2$ does *not* go to 0 as $n\to\infty$.

A moment's reflection reveals that the inconsistency argument is ill-thought out. The is because the incidental parameter problem renders $\widehat{\mu}_t{=}\frac{1}{2}(X_{1t}+X_{2t})$ *an inconsistent* estimator of μ_t, for $t{=}1,2,...,n$; there are only two observations for each $\widehat{\mu}_t$, and thus their variance $Var(\widehat{\mu}_t){=}\frac{1}{2}\sigma^2$ does not go to zero as $n\rightarrow\infty$.

What the critics of the ML method do not appreciate enough is the fact that treating the unknown parameters $(\mu_1,\mu_2,...,\mu_n)$ as incidental and designating σ^2 the only parameter of interest, does not let the statistician 'off the hook'. This is because the parameter of interest $\sigma^2{=}E(X_{it}{-}\mu_t)^2$, defining the variation around μ_t, invokes the incidental parameters. Put more intuitively, when the data come in the form of $\mathbf{Z}_0{:=}(\mathbf{z}_1,\mathbf{z}_2,...,\mathbf{z}_n)$, $\mathbf{z}_t{:=}(x_{1t},x_{2t})$, one can get to σ^2 via μ_t, and using $\widehat{\mu}_t$ leads to problems because it is an inconsistent estimator. In that sense (Severini, 2000):

"... this model falls outside of the general framework we are considering since the dimension of the parameter $(\mu_1,\mu_2,...,\mu_n,\sigma^2)$ depends on the sample size." (p. 108)

That is, calling $\widehat{\sigma}^2{=}\frac{1}{2n}\sum_{t=1}^{n}[(X_{1t}{-}\widehat{\mu}_t)^2{+}(X_{2t}{-}\widehat{\mu}_t)^2]$ a MLE is highly misleading since the ML method was never meant to be applied to statistical models whose number of unknown parameters increases with the sample size n.

In truth, one should be very skeptical of any method of estimation which yields consistent estimators in cases where the statistical model in question is *ill-defined*, as in the case of the incidental parameter problem. Hence, the more interesting question should be:

why would the ML method yield a consistent estimator of σ^2?

The fact that the ML method does *not* yield a consistent estimator of σ^2 should count in its favor not against it! To paraphrase Stigler's quotation: the ML method 'warns the modeler that the information in the data has been spread too thinly'.

4 Recasting the original Neyman-Scott model

The question that naturally arises is: can one respecify the above statistical model to render it well-defined but retaining the parameter of interest? The answer is surprisingly straightforward. Since the incidental parameter problem arises because of the unknown but *t-varying* means $(\mu_1,\mu_2,...,\mu_n)$, one can re-specify the original bivariate model into a univariate *simple Normal* (one parameter), using the transformation:

$$\boxed{Y_t{=}\tfrac{1}{\sqrt{2}}(X_{1t}-X_{2t})\backsim \mathsf{NIID}\left(0,\sigma^2\right),\ t{=}1,2,...,n,...,}$$

$$E(Y_t){=}\tfrac{1}{\sqrt{2}}E(X_{1t}-X_{2t}){=}\tfrac{1}{\sqrt{2}}(\mu_t-\mu_t){=}0,$$

$$Var(Y_t)=\tfrac{1}{2}[Var(X_{1t})+Var(X_{2t})]{=}\sigma^2.$$

This is a sensible thing to do because taking the difference eliminates the *nuisance* parameters $(\mu_1,\mu_2,...,\mu_n)$, without affecting the parameter of interest.

For this simple Normal model, the MLE for σ^2 is: $\widehat{\sigma}^2_{MLE}=\frac{1}{n}\sum_{t=1}^{n}Y_t^2$,

which it is *unbiased, fully efficient* and *strongly consistent:*

$$E(\widehat{\sigma}^2_{MLE})=\sigma^2,\ Var(\widehat{\sigma}^2_{MLE})=\frac{2\sigma^4}{n},\ \widehat{\sigma}^2_{MLE}\overset{a.s.}{\to}\sigma^2.$$

Notes:

(i) The above recasting of the Neyman-Scott model can be easily extended to the case $(X_{1t}, X_{2t}, ..., X_{mt})\,,\ 2\leq m<n.$

(ii) Hald (2007), p. 182-3 offers an alternative, highly original, way to sidestep the incidental parameter problem using Fisher's two stage ML method.

5 Conclusion

The main conclusion from the above discussion is that when the ML method gives rise to inconsistent estimators, the modeler should take a closer look at the assumed statistical model; chances are, it is ill-defined. This is particularly true in the case where the assumed model suffers from the incidental parameter problem. In such cases the way forward is to recast the original model to render it well-defined and then apply the ML method. This argument is illustrated above using the Neyman-Scott (1948) model.

References

[1] Cox, D. R. (2006), *Principles of Statistical Inference*, Cambridge University Press, Cambridge.

[2] Cramer, H. (1946), *Mathematical Methods of Statistics*, Princeton University Press, Princeton, NJ.

[3] Fisher, R. A. (1922), "On the mathematical foundations of theoretical statistics", *Philosophical Transactions of the Royal Society A*, **222**: 309-368.

[4] Hald, A. (2007), *A History of Parametric Statistical Inference from Bernoulli to Fisher, 1713-1935*, Springer, NY.

[5] Le Cam, L. (1953), "On some asymptotic properties of maximum likelihood estimates and related Bayesian estimates," *University of California Publications in Statistics*, 1: 277-330.

[6] Neyman, J. and E. L. Scott (1948), "Consistent estimates based on partially consistent observations," *Econometrica*, 16: 1-32.

[7] Severini, T. A. (2000), *Likelihood Methods in Statistics*, Oxford University Press, Oxford.

[8] Stigler, S. M. (2007), "The Epic Story of Maximum Likelihood," *Statistical Science*, 22: 598-620.

[9] Wald, A. (1949), "Note on the consistency of the maximum likelihood estimate," *Annals of Mathematical Statistics*, 20: 595-601.